View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

PERGAMON

Vision Research 42 (2002) 2617–2634

Vision Research

www.elsevier.com/locate/visres

A spectral histogram model for texton modeling and texture discrimination

Xiuwen Liu [a,*], DeLiang Wang [b]

[a] Department of Computer Science, Florida State University, Tallahassee, FL 32306-4530, USA

[b] Department of Computer and Information Science, Center for Cognitive Science, The Ohio State University, 2015 Neil Avenue, Columbus, OH 43210, USA

Received 14 August 2001; received in revised form 9 April 2002

Abstract

We suggest a spectral histogram, defined as the marginal distribution of filter responses, as a quantitative definition for a texton pattern. By matching spectral histograms, an arbitrary image can be transformed to an image with similar textons to the observed. We use the χ^2-statistic to measure the difference between two spectral histograms, which leads to a texture discrimination model. The performance of the model well matches psychophysical results on a systematic set of texture discrimination data and it exhibits the nonlinearity and asymmetry phenomena in human texture discrimination. A quantitative comparison with the Malik–Perona model is given, and a number of issues regarding the model are discussed.
© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Texton modeling; Texture discrimination; Texture Synthesis; Texture perception

1. Introduction

Texture perception is one of the pillars in the study of early visual perception (Beck, 1966; Julesz, 1962). Much of the psychophysical work concentrates on texture discrimination, or detecting whether two texture patches can be discerned rapidly by human observers (for reviews see Bergen, 1991; Papathomas, Chubb, Gorea, & Kowle, 1995). Effortless texture discrimination takes place rapidly and is viewed as a preattentive process that occurs in parallel across the whole visual field. A critical empirical issue is what stimulus conditions result in preattentive texture segregation as opposed to a slow, effortful process that requires focal attention. Many texture patterns have been devised to test various ideas and hypotheses on this issue, and have revealed an array of perceptual phenomena concerning texture discrimination.

Beck, a pioneer in texture perception, described a multistage conceptual model for texture segregation in 1982. According to his model (Beck, 1982), the first stage performs local feature detection with receptive fields in the visual system. The second stage extracts the total differences in color, luminance, orientation, and size between neighboring texture elements. The last stage segregates an image into regions of the same texture on the basis of the magnitude and distribution of difference signals.

In a life-long effort to pursue a scientific theory for texture perception similar to that of the Young–Helmholtz trichromatic theory for color perception, Julesz and his colleagues are the most influential in conceptual thinking about texture perception as well as in setting the empirical agenda on the investigation of texture discrimination. After extensive formulations and reformulations in terms of high-order statistics, Julesz eventually proposed the texton theory for texture perception. According to the texton theory, textures are discriminated if they differ in the density of certain simple, local textural features, or textons (Julesz, 1981, 1995). Three textons have been consistently specified (Julesz, 1981, 1986): elongated blobs defined by color, orientation, size, etc., line terminators, and line crossings. Collinearity and local closure are often mentioned in the literature as well. Though theorized by Julesz as

* Corresponding author. Tel.: +1-850-644-0050; fax: +1-850-644-0058.

E-mail addresses: liux@cs.fsu.edu (X. Liu), dwang@cis.ohio-state.edu (D. Wang).

0042-6989/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0042-6989(02)00297-3

perceptual atoms, "What textons really are is hard to define" (Julesz, 1995, p. 134). As pointed out by Bergen and Adelson, one major difficulty with his approach is that "it is based on a verbal description of image features" (Bergen & Adelson, 1988, p. 363).

More recently, texture discrimination has received considerable attention from the computational perspective, and many models have been proposed (see among others Barth, Zetzsche, & Rentschler, 1998; Caelli, 1985; Fogel & Sagi, 1989; Graham, Beck, & Sutter, 1992; Malik & Perona, 1990; Räth & Morfill, 1997; Rubenstein & Sagi, 1990; Turner, 1986; Voorhees & Poggio, 1988). Although these models differ in details, they all share a three-stage common structure, referred to as filter–rectify–filter (FRF) model by Wilson (1999). The first stage is filtering by a set of linear filters. The second stage is a nonlinear rectification process, followed by some pooling operation (lowpass filtering). Rectification removes or inverts negative filter responses, and compresses the range of responses. Note that the rectifying nonlinearity is necessary, since, otherwise, linear filters will give responses that cannot discriminate in the third stage two texture patches with the same mean luminance. The pooling operation performs some spatial averaging, smoothing, or nonlinear inhibition to remove inhomogeneity in rectified filter responses. This is necessary because filters are regularly laid out, whereas texture elements on an image are not. Without pooling filter responses within a homogeneous texture region would produce inhomogeneous responses due to the misalignment between texture elements and filters, creating problems for subsequent processing. The third stage—the second filtering stage—determines texture boundary, or equivalently, produces texture regions on the basis of some edge/contour detection. We note that this common structure is very similar to Beck's conceptual model described earlier.

The above computational models are mainly motivated by perceptual data obtained from synthetic textures, which are easy to modify in a controlled way—often necessary for collecting systematic psychophysical data. Moreover, these models mainly aim at simulating perceptual data or exhibiting perceptual phenomena. Though some of these models are quite successful in accounting for empirical data on texture discrimination, they do not provide an explicit model for texton or texture itself. By an explicit model we mean that it gives an explicit description of what a texture is. The above models aim at producing boundaries between different textures, which do not necessarily lead to texture models beyond boundary detection, and their adequacy as a texture model is never checked. Thus, these quantitative efforts do not seem to provide much insight into the nature of texture perception in Julesz's sense.

On the other hand, explicit texture modeling has been undertaken in computer vision. The most popular is a class of Markov random field (MRF) models (Cross & Jain, 1983; Dubes & Jain, 1993; Geman & Geman, 1984). MRF models define probability distributions based on some correlation within a local neighborhood. Once an MRF model and its parameters are specified, a statistical sampler can be used to generate (synthesize) individual textures that realize the model. Texture models of this kind have been successfully applied to texture image processing, including classification and segmentation. On the other hand, these models cannot account for the kind of textures often used in psychophysical experiments; such textures, though mostly synthetic and binary, generally have particular elements and global structure, and as a result they are very difficult to describe and synthesize on the basis of local statistical models.

Motivated by the observation that two textures are often difficult to discriminate if they trigger filter responses with a similar distribution, Heeger and Bergen (1995) proposed a histogram-based model for texture synthesis. Given an original texture, their model first decomposes it into an image pyramid using a bank of steerable filters (Simoncelli, Freeman, Adelson, & Heeger, 1992) that are sensitive to both orientation and scale of the image structure. For each subband image in the pyramid, its response histogram is calculated. During the synthesis stage, their algorithm attempts to transform an initial noise image into a similar texture using the same filter bank by applying the following procedure iteratively. At each iteration, the current synthesized image is first decomposed into an image pyramid as for the original texture. Then each subband image is transformed using a deterministic histogram matching algorithm so that its histogram matches the corresponding one in the original pyramid. An updated image is subsequently generated by inverting the transformed subband images, which is a computationally advantageous property of the pyramid representation. They have reported impressive results for natural textures, and the study has motivated considerable subsequent research into texture synthesis (Portilla & Simoncelli, 2000; Zhu, Liu, & Wu, 2000; Zhu, Wu, & Mumford, 1997). Though not explicitly stated, the Heeger and Bergen study implies a texture model that corresponds to the histograms of an image pyramid. To our knowledge, however, no histogram-based model has been used to address human texture discrimination.

In this paper, we study a version of the histogram-based model, called spectral histogram, for simulating human texture discrimination; in particular, we suggest a spectral histogram as a quantitative definition for a texton pattern. The spectral histogram model consists of marginal distributions of responses from a bank of filters within an image window. We show that this model elegantly avoids both the rectifying nonlinearity and subsequent pooling in FRF models, thus resulting in a

more parsimonious model. The adequacy of the model is established by extensive results on synthesizing both synthetic and natural textures using an effective sampling algorithm. To address texture discrimination, we employ the χ^2-statistic to measure the distance between two spectral histograms. This model yields surprisingly good performance on a systematic set of texture discrimination data. This performance is compared with that of the Malik and Perona model (Malik & Perona, 1990). The spectral histogram model demonstrates the nonlinearity of human texture discrimination. Furthermore, we illustrate that it can exhibit the asymmetry phenomenon in texture discrimination.

This paper is organized as follows. Section 2 describes the spectral histogram model and verifies its adequacy as a texton model by synthesizing texture patterns with distinct spectral histograms. Section 3 simulates a set of psychophysical data on texture discrimination, and draws a comparison with the Malik–Perona model. Section 4 relates the special histogram model with other studies on texture modeling and discrimination. Section 5 discusses a number of issues including biological plausibility. Section 6 concludes the paper.

2. Spectral histogram for texton modeling

Based on psychophysical and neurophysiological data, it is widely accepted that the human visual system transforms a retinal image into a local spatial/frequency representation. Such a representation can be simulated by a bank of filters with tuned frequencies and orientations, e.g. Gabor filters, and finds applications in many areas including image compression. For texture modeling, filter responses themselves are not adequate as textures are regional properties, as demonstrated by a recent comprehensive study on filter-response methods for texture classification (Randen & Husoy, 1999). The result shows that all the methods included in the study fail to produce meaningful classification results for a set of textures, suggesting that filter responses are not sufficient to characterize textures.

Within the spatial/frequency representation, additional steps seem necessary in order to address the inadequacy of filter responses. One reasonable step would be to integrate information from filter responses so as to form perceptually meaningful feature statistics for textures. Studies of human texture perception (Bergen & Adelson, 1988; Chubb, Econopouly, & Landy, 1994) show that two textures are often perceptually similar when they give a similar distribution of responses from a bank of filters. A recent study (Kingdom, Hayes, & Field, 2001) demonstrates that human observers are sensitive to histogram differences in synthetic wavelet-textures.

Motivated by perceptual observations and the Heeger and Bergen texture synthesis model (Heeger & Bergen, 1995), we describe a *spectral histogram* model within the local spatial/frequency representation framework, for characterizing a texton pattern. We then apply the model to texture discrimination in the next section.

2.1. Definition and properties

Given an input image window $\mathbf{W}$ and a bank of filters $\{F^{(\alpha)}, \alpha = 1, 2, \ldots, K\}$, we compute, for each filter $F^{(\alpha)}$, a subband image $\mathbf{W}^{(\alpha)}$ through linear convolution, i.e., $\mathbf{W}^{(\alpha)}(v) = F^{(\alpha)} * \mathbf{W}(v) = \sum_u F^{(\alpha)}(u)\mathbf{W}(v-u)$, whereby a circular boundary condition is used for convenience. For $\mathbf{W}^{(\alpha)}$, we define the histogram as $H^{(\alpha)}_{\mathbf{W}}(z) = 1/|\mathbf{W}| \sum_v \delta(z - \mathbf{W}^{(\alpha)}(v))$, which corresponds to the marginal distribution. [1] We then define the spectral histogram with respect to the chosen filters as

$$H_{\mathbf{W}} = (H^{(1)}_{\mathbf{W}}, H^{(2)}_{\mathbf{W}}, \ldots, H^{(K)}_{\mathbf{W}}). \tag{1}$$

This definition reflects the assumption that a texture is defined collectively by responses of different filters. According to (1), the spectral histogram of an image or an image patch is essentially a vector consisting of marginal distributions of filter responses. The size of the input image window, $|\mathbf{W}|$, is called the *integration scale*. Because the marginal distribution of each filter response is a probability distribution, we define a similarity measure as χ^2-statistic, which is used widely to compare two histograms,

$$\chi^2(H_{\mathbf{W}_1}, H_{\mathbf{W}_2}) = \frac{1}{K}\sum_{\alpha=1}^{K}\sum_{z}\frac{(H^{(\alpha)}_{\mathbf{W}_1}(z) - H^{(\alpha)}_{\mathbf{W}_2}(z))^2}{H^{(\alpha)}_{\mathbf{W}_1}(z) + H^{(\alpha)}_{\mathbf{W}_2}(z)}. \tag{2}$$

The spectral histogram integrates responses from different filters and provides a naturally normalized feature statistic to compare images of different sizes. By implicitly integrating geometrical and photometric structures of textures, the spectral histogram provides a sufficient model for characterizing perceptual appearance of textures (Liu, 1999).

Since a bin in a histogram counts how many of the identical filters generate a similar response within a spatial window that is substantially larger than the size of a texture element, a spectral histogram is fundamentally insensitive to precise locations of texture elements within the window. This property is illustrated in Fig. 1. Fig. 1a and b show two textures with similar spectral histograms, and thus the images would belong

[1] In statistical modeling of images, by associating each pixel with a random variable, $\mathbf{W}^{(\alpha)}$ is viewed as one sample from the underlying joint distribution. Under the assumption that the random variables are independent and identically distributed, the joint distribution is completely determined by the marginal distribution and the histogram is the maximum likelihood estimate of the marginal (Duda, Hart, & Stork, 2000) and thus the underlying distribution.

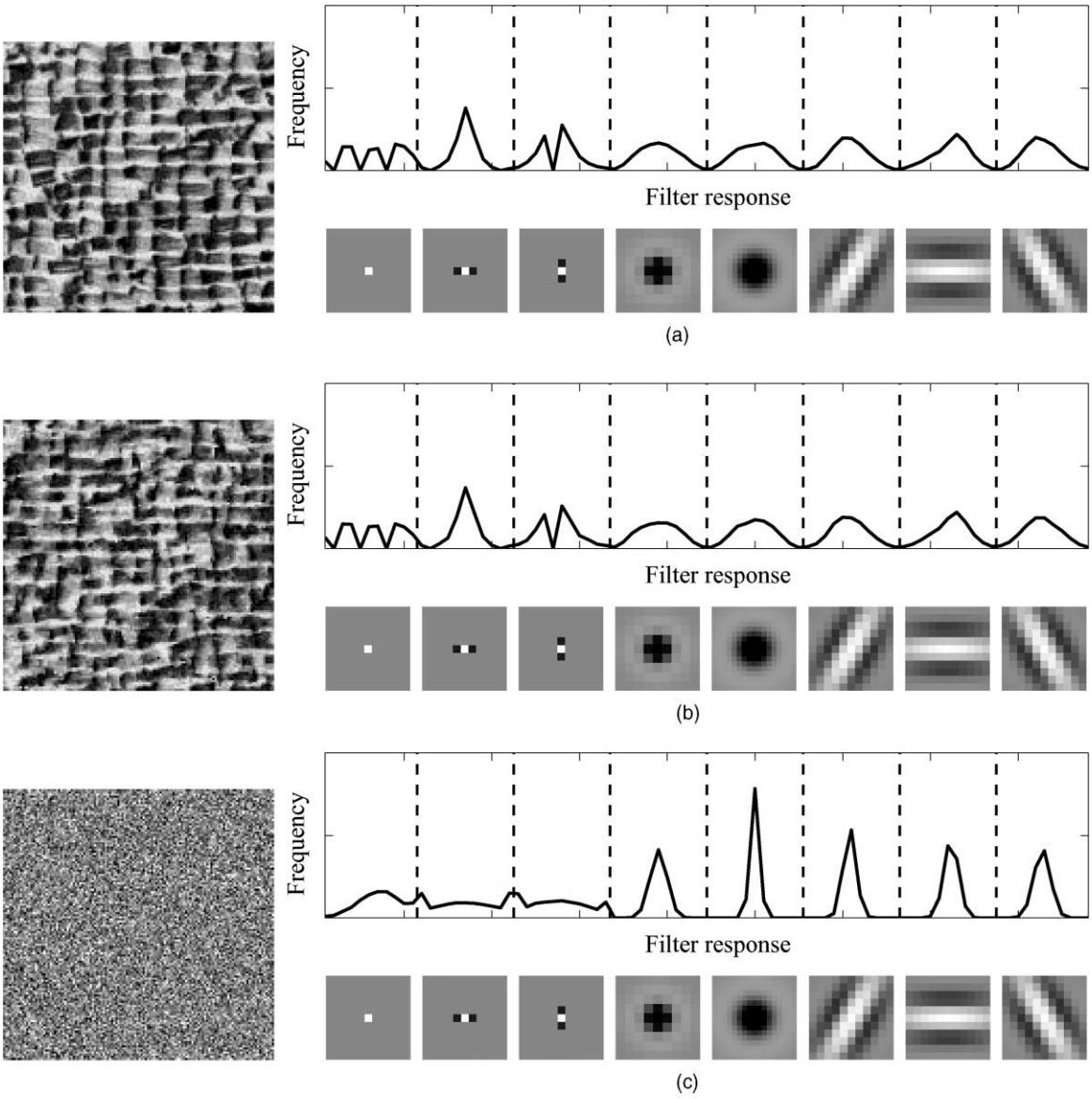

Fig. 1. Patches with similar histograms that are perceptually indiscriminable and those with dissimilar histograms that are perceptually different. Here eight filters, consisting of the intensity filter, two local difference filters, two LoG filters, and three Gabor filters, are used to calculate the spectral histogram, and their corresponding histograms are separated by dash lines with filter profiles shown below. Here profiles are scaled for illustration purposes. The size of all the images is 128×128 and pixel values are between 0 and 255. (a, b) Two patches with their corresponding spectral histograms. The spectral histograms are similar. However, the root-mean-square distance between the two patches is large—94.0 per pixel. (c) A Gaussian noise image with its spectral histogram. The root-mean-square distance between this patch and that in (a) is 84.5 per pixel, smaller than the distance between (a) and (b).

to the same texture. Pixelwise, however, the two images are very different. For example, the root-mean-square distance between them is larger than that between Fig. 1a and c, the latter being a Gaussian noise image. We emphasize that this important characteristic is consistent with the evidence from human texture discrimination that "only the number (or density) of textons has perceptual significance and their position is ignored" (Julesz, 1981, p. 97).

Formally, the spectral histogram model exhibits desired properties for texton modeling. Because filter responses depend only on relative locations of pixels, the exact position of the image window $\mathbf{W}$ does not affect its spectral histogram as long as it encloses the same texture region, thus resulting in translation invariance. Because the histogram function is nonlinear due to the usage of the delta function, the spectral histogram is also nonlinear. To see this, let $\mathbf{W}$ be a nonzero uniform image window, i.e., $\mathbf{W}(u) = c$ for all u, where c is a nonzero constant. Let $\mathbf{W}_1 = \beta\mathbf{W}$ and $\mathbf{W}_2 = (1-\beta)\mathbf{W}$, where $0 < \beta < 1$, and thus $\mathbf{W}_1 + \mathbf{W}_2 = \mathbf{W}$. For a given $F^{(\alpha)}$, let $\mathbf{W}^{(\alpha)}(v) = F^{(\alpha)} * \mathbf{W}(v) = \sum_u F^{(\alpha)}(u)c = c_1$ for all v.

Because of the linear convolution, we have $\mathbf{W}_1^{(\alpha)}(v) = \beta c_1$ and $\mathbf{W}_2^{(\alpha)}(v) = (1-\beta)c_1$. Thus we have $H_{\mathbf{W}}^{(\alpha)}(z) = \delta(z - c_1)$, $H_{\mathbf{W}_1}^{(\alpha)}(z) = \delta(z - \beta c_1)$, and $H_{\mathbf{W}_2}^{(\alpha)}(z) = \delta(z - (1-\beta)c_1)$. For all $F^{(\alpha)}$ where c_1 is not zero, we have $H_{\mathbf{W}}^{(\alpha)} \neq H_{\mathbf{W}_1}^{(\alpha)} + H_{\mathbf{W}_2}^{(\alpha)}$. [2] In addition, multiple filters impose different constraints on the geometrical structures of images within the same spectral histogram, and this makes linear summation not applicable to spectral histograms. The nonlinearity of human texture discrimination was demonstrated by Williams and Julesz (1992a), and was used as evidence against any linear model (e.g. Bergen & Adelson, 1988). The issue of nonlinearity will be further discussed in Section 3.

We have shown elsewhere that a spectral histogram can uniquely represent an image up to a translation given sufficient filters (Liu & Wang, 2000). Intuitively, each filter provides a constraint on the set of images that share the spectral histogram of the image. By adding more and more filters, the set becomes more and more constrained and it can be eventually made to contain only the image and its translations. Also, with appropriately chosen filters the spectral histogram provides a unified texture feature statistic, where many existing texture features can be treated as special cases (see Liu, 1999).

2.2. Texton patterns as spectral histograms

Besides the problem of being a verbal description (Bergen & Adelson, 1988), the notion of textons as conspicuous local features implies that textons are perceptual properties. This seems at odds with the evidence that texture segregation takes place at a level earlier than the one at which perceptual features can be derived (Bergen, 1991). Even for visual cortical cells with Gabor-like receptive fields, which are frequently taken as edge or line detectors, they respond also to sinusoidal gratings, white noise, and many other patterns. Textons such as corner and closure detectors are more specialized and complex to compute, and thus would presumably arise even later in the visual processing pathway.

We suggest a filter histogram as a quantitative definition for a texton pattern. A texton, according to this suggestion, would simply correspond to a filter. The entire spectral histogram given in (1), which consists of multiple filter histograms, defines a texture. The computation leading to a spectral histogram involves commonly used spatial/frequency filters, and thus our definition does not invoke perceptual attributes. Our definition is primarily based on the observation that texture images with a similar histogram are composed of similar elements and similar densities; as such, they would appear perceptually similar, as shown in the next subsection.

2.3. Texture synthesis

To verify the sufficiency of the spectral histogram model we have performed extensive texture synthesis experiments. Given an observed texture, such as the one shown in Fig. 2a, we compute its spectral histogram, which encodes the perceptual structure of the image implicitly. To check the sufficiency of the spectral histogram for characterizing textures, we then generate images that satisfy the constraints $H_{\mathbf{I}} = H_{\text{obs}}$, where $\mathbf{I}$ is an image, $H_{\mathbf{I}}$ its spectral histogram, and H_{obs} the spectral histogram of the observed image.

In the following simulations we use a fixed set of 47 filters; these are two local difference filters: $D_{xx} = [-1.0 \quad 2.0 \quad -1.0]$ and $D_{yy} = [-1.0 \quad 2.0 \quad -1.0]^t$ (one along a row and one along a column with t indicating transpose), three Laplacian of Gaussian (LoG) filters: $\text{LoG}(x, y|T) = (x^2 + y^2 - T^2)\exp\{-(x^2 + y^2)/T^2\}$ (with T set to $\sqrt{2}/2$, 1, and 2), and 42 Gabor filters: $\text{Gabor}(x, y|T, \theta) = \exp\{-(1/2T^2)(4(x\cos\theta + y\sin\theta)^2 + (-x\sin\theta + y\cos\theta)^2)\}\cos((2\pi/T)(x\cos\theta + y\sin\theta))$ (with T set to 2, 4, 6, 8, 10, 12 and 14 and six equally-spaced orientations at each scale) to characterize texton patterns. Note that the specific forms of filters are not critical for the spectral histogram representation (see Fig. 9 for example).

Due to the high dimensionality of $\mathbf{I}$ (for a 128×128 image, the dimension is 16 384), the constraints of a spectral histogram need to be satisfied through stochastic simulation because traditional deterministic search methods are computationally not feasible. One commonly used method is the Gibbs sampler (Geman & Geman, 1984), which has been demonstrated to be effective for natural textures (Zhu et al., 2000). Essentially the Gibbs sampler tries to reduce the error between the given histograms and the ones of the current image following a statistical procedure. In the binary image case, it computes the errors using the black and white intensity at a pixel location and the resulting new value is set with a higher probability to the one with the smaller error. The probability is also controlled by a gradually reduced temperature parameter. The constraints of different filters are incorporated in the error evaluation between histograms. In practice, the effectiveness of the sampler critically depends on the temperature parameter and can be easily trapped at local minima (i.e. suboptimal results); Fig. 2b shows a typical example of such failure. Our experiments show that the problem becomes worse for gray-level textures.

To explore the image space more effectively, we utilize a sampling procedure similar to that given by Zhu et al. (1997). The procedure was originally proposed to learn parameters in a probability model. Here it is used as a

[2] A normalization is needed when two spectral histograms are summed together.

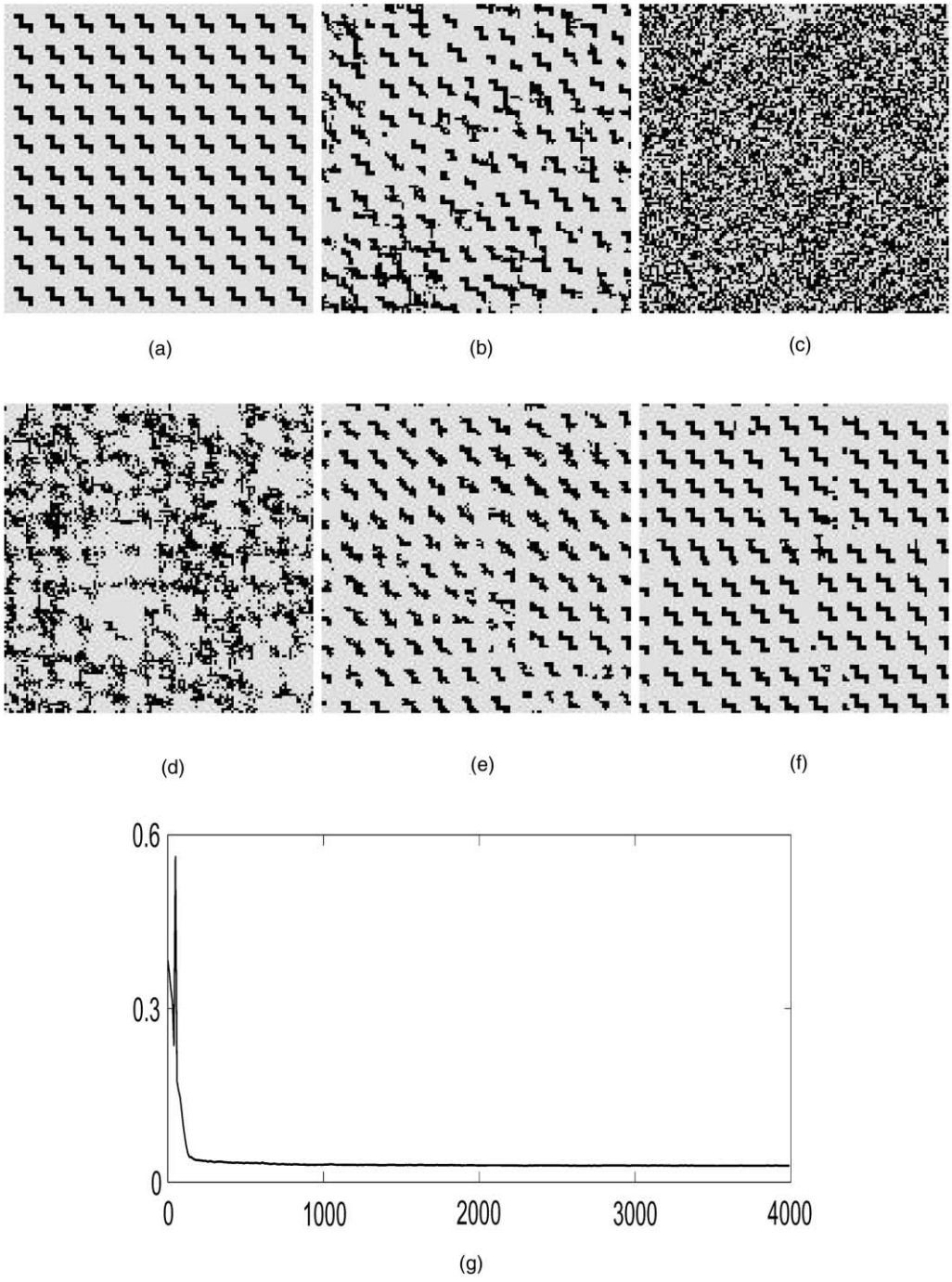

Fig. 2. A texture and synthesized images at different sweeps. The size of the image is 128×128. (a) Observed image. (b) A synthesized image using the Gibbs sampler. The error per filter, defined as $\sum_{\alpha=1}^{K} \sum_{i=1}^{L^{(\alpha)}} |H_{\mathbf{I}_{\text{syn}}}^{(\alpha)}(i) - H_{\text{obs}}^{(\alpha)}(i)|/K$, is 0.116. (c) Initial image for sampling. (d)–(f) Synthesized images at sweep 40, 100, and 4000 with the error per filter of 0.237, 0.098, and 0.028 respectively. (g) The error per filter with respect to the number of sweeps.

sampling algorithm to effectively explore the space in which the spectral histogram matches the observed one. By updating parameters along the sampling process, the resulting algorithm eliminates the temperature parameter. A version of the sampling algorithm for binary textures is given in Appendix A. Fig. 2c shows the initial condition for the sampling procedure, which is a white noise image. Fig. 2d and e show intermediate images at sweep 40 and 100 respectively. Fig. 2f shows the synthesized image at sweep 4000, which is perceptually similar to the observed. The texture element is synthesized very well through the constraints imposed by different filters; the global structure is also reproduced. Fig. 2g shows the average histogram error per filter with respect to the number of sweeps. As is evident from Fig. 2g, there exist local minimum states, and our sampling

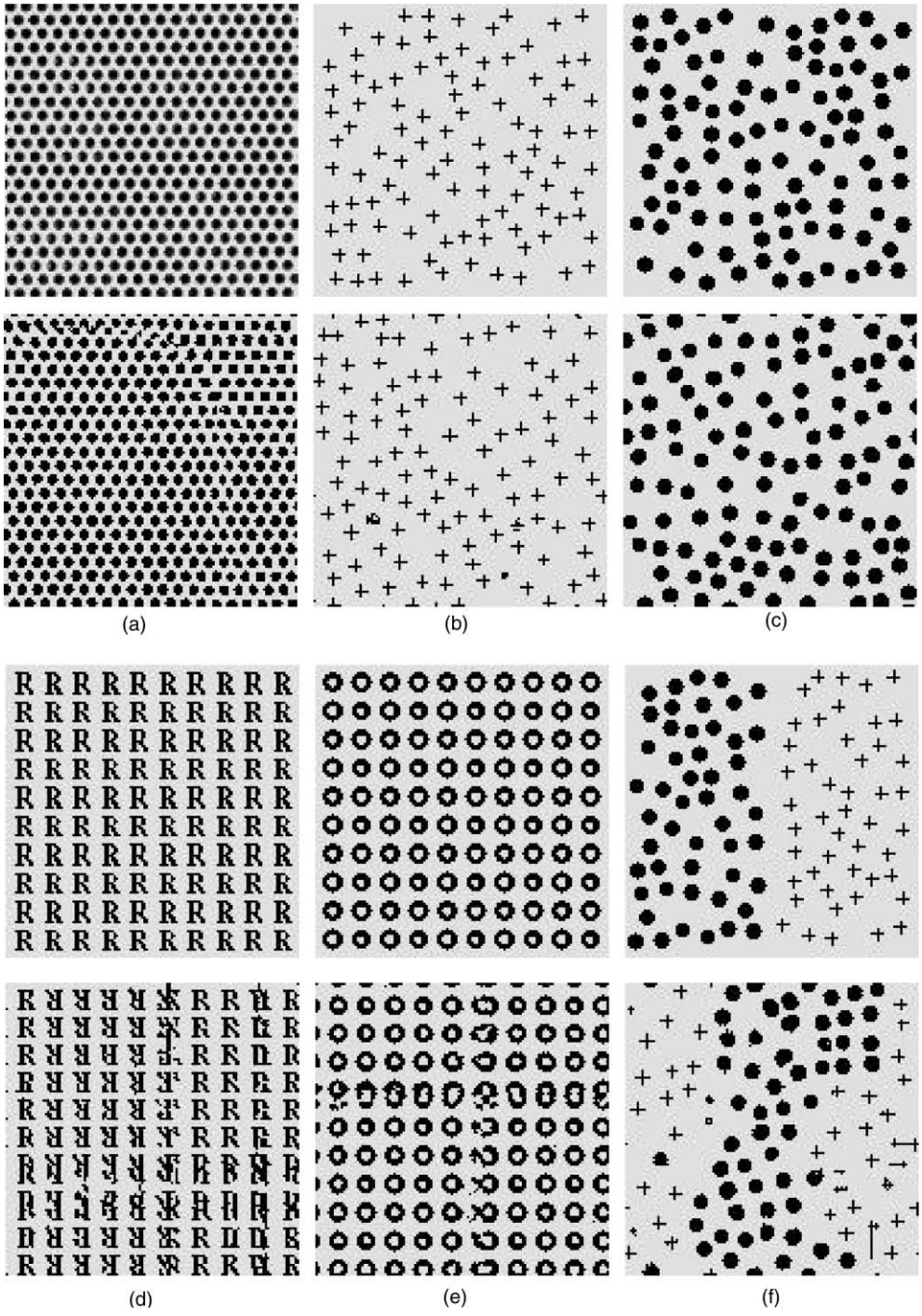

Fig. 3. Synthesized images for synthetic textures with different micropatterns. In each column, the upper part shows the observed texture and the lower part a synthesized texture at sweep 4000. (a) A texture consisting of regularly arranged hexagons. (b) A texture consisting of pluses. (c) A texture with filled circles. (d) A texture consisting of R's. (e) A texture consisting of empty circles. (f) An image consisting of two distinct textures.

procedure overcomes local minima and reaches a globally optimal state.

Fig. 3 shows several more examples. Fig. 3a shows a synthesized texture consisting of hexagons arranged regularly. The spectral histogram captures the hexagons as well as the regular structure. Fig. 3b shows a texture consisting of randomly placed pluses and Fig. 3c circles. The micropatterns are captured by their spectral histograms. Fig. 3d shows that 'R' can be reproduced using the spectral histogram. Fig. 3e shows a texture of empty circles placed on a regular grid. The structure of each element is synthesized solely based on the spectral histogram. Worth noting in the above examples is that the regular layout of texture elements is very well captured

by the spectral histogram model. Fig. 3f shows an interesting case, where one spectral histogram captures both circle and plus elements at the same time (recall that boundary wrap-around is employed).

Note that the filters are fixed for all the synthesis examples and there is no explicit template for a texture element. The basic elements are captured by the spectral histograms through imposed constraints by different filters. This offers distinct advantages over texture models based on explicit templates (Voorhees & Poggio, 1988). Not only must a large number of templates be specified to model different kinds of textures, but also must the elements appeared in the observed image be extracted, which is computationally expensive. In addition, a perceptual distance between textures still needs to be defined as textures consisting of different templates need to be compared for discrimination (see Fig. 5 for example).

The spectral histogram is perceptually sufficient not only for synthetic texture patterns, but also for natural images, as shown by Heeger and Bergen (1995), Zhu et al. (1997), Zhu et al. (2000), and Liu (1999). For example, Fig. 4a shows a cheetah image and Fig. 4b shows a patch of cheetah skin. Fig. 4c shows the synthesized patch by matching the spectral histograms. The synthesized image captures the perceptual characteristics of the cheetah skin.

The above results clearly demonstrate that different images with similar spectral histograms yield perceptually similar appearances. These results on synthetic images, together with extensive results on natural textures, suggest that spectral histograms capture a level of image description that is sensitive to certain types of spatial information such as orientation, scale, and density, while oblivious to elaborate geometrical properties. A texture model requires a balance between descriptions that are too simple to reveal anything different and those that are too complex to generate any abstraction of an image (Watt, 1995). The spectral histogram model, we believe, strikes a balance of this kind.

3. Texture discrimination

The previous section demonstrates that the spectral histogram model provides a viable definition for textures. Given that much of psychophysical data on texture perception is on comparing texture images, a critical evaluation of any attempt for quantitative texton modeling is to match psychophysical data of texture

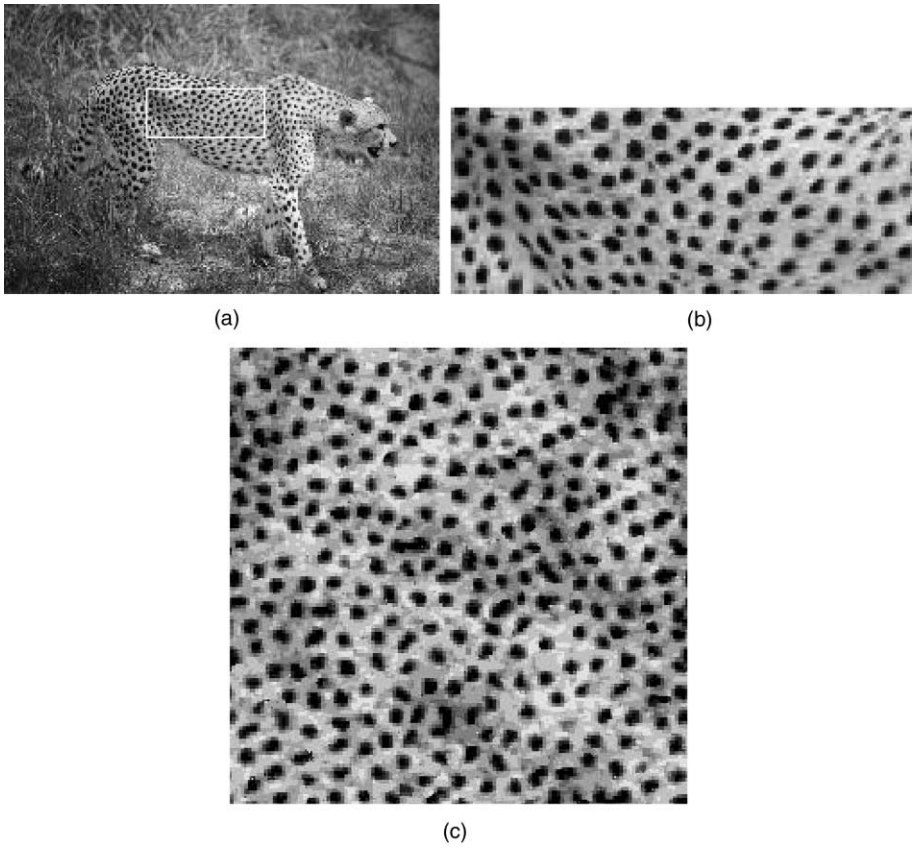

Fig. 4. Natural texture of cheetah skin. (a) An image containing a cheetah. The size of the image is 648×972. (b) The cheetah skin from the enclosed area in (a). The size of this area is 104×258. (c) A synthesized image of 256×256.

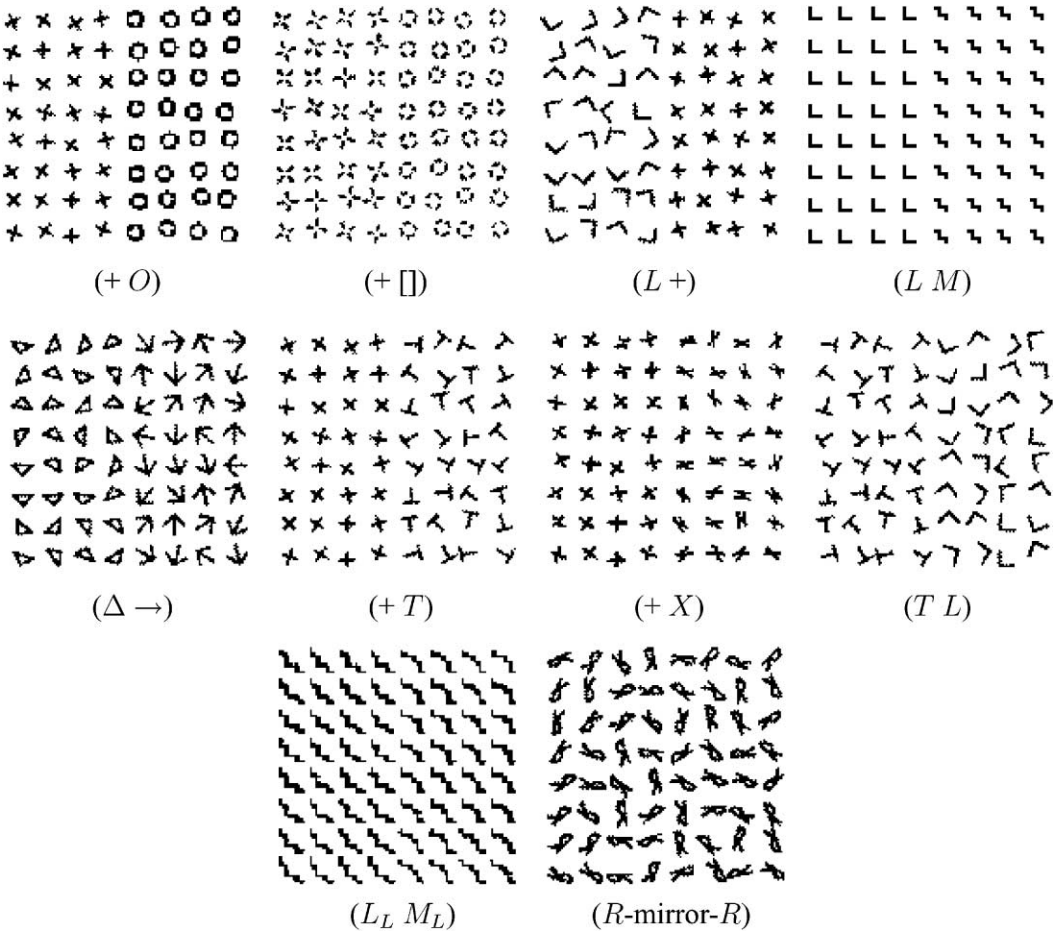

Fig. 5. Ten texture pairs scanned from Malik and Perona (1990). The size of all the scanned images is 154 × 154.

discrimination. This section tests our model with a set of systematic human data on texture discrimination. The set consists of 10 texture pairs, as shown in Fig. 5. Seven are from Kröse (1986), two from Williams and Julesz (1992a), and one composed of R's and mirror-image R's (called R-mirror-R). The same 10 texture pairs were used to evaluate the well-known Malik and Perona model (Malik & Perona, 1990), thus facilitating a quantitative comparison with their model. The texture pairs shown in Fig. 5 were scanned from Malik and Perona (1990).

Eqs. (1) and (2) essentially constitute our model for texture discrimination. We adopt similar procedures used by Malik and Perona (1990) for testing texture discrimination performance. Instead of using 96 pairs of filters in Malik and Perona (1990), we use the same two gradient filters and three LoG filters used in the synthesis experiments. Gabor filters are not used for discrimination because orientation is not a major factor for discriminating the texture pairs in Fig. 5. At each pixel location, we extract local spectral histograms at integration scale 29×29, i.e. over a window of 29×29 pixels centered at the location, and the gradient is the average χ^2-distance per filter between the spectral histograms of the two adjacent windows along a row. Then the gradient is averaged along each column as done in Malik and Perona (1990). The texture gradients generated from our method for the two texture pairs (+ O) and (R-mirror-R) are shown in Fig. 6b and d.

Several observations can be made from the gradient results of Fig. 6. First, a texture pattern does not give rise to a homogeneous texture region, and variations within each texture region are clearly visible. For regularly arranged micropatterns people do perceive distinct columns besides the middle boundary that separates two main texture regions; see the texture pair (+ O) in Fig. 5. Second, because of the variations among different micropatterns, the absolute value of texture gradient should not be used directly as a measure for texture discrimination as in Malik and Perona (1990). As shown in Fig. 6d, even though the gradient is much weaker compared to Fig. 6b, the filters still respond to element variations, which is also evident in Malik and Perona (1990). However, no texture boundary is perceived in this case.

Based on these observations, we propose a texture discrimination measure as the difference between the central peak and the maximum of two adjacent side

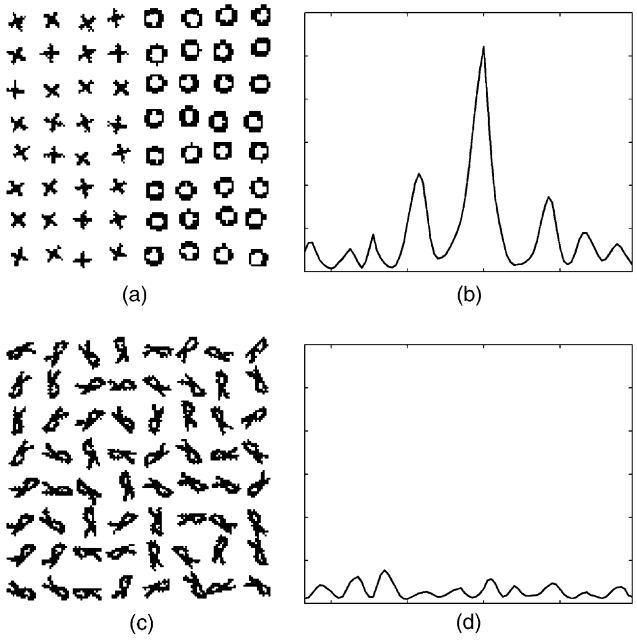

Fig. 6. The averaged texture gradient for two selected texture pairs in Fig. 5. (a) The texture pair (+ O). (b) The texture gradient averaged along each column for (a). The horizontal axis is the column number and the vertical axis is the gradient. (c) The texture pair (R-mirror-R). (d) The texture gradient for (c).

peaks. In other words, the peak corresponding to the middle boundary is compared against the two adjacent ones corresponding to the interior boundaries within each texture. In the (+ O) case, the central peak is 0.239, and the left and right side peaks are 0.104 and 0.08 respectively. Thus the discrimination measure is 0.135. For the (R-mirror-R) case, the central peak is 0.017 and the left and right side peaks are 0.012 and 0.027 respectively. Thus the measure is −0.01, indicating that the two texture regions are not discriminable at all.

We calculate the proposed discrimination measure for the 10 texture pairs. Table 1 shows our results along with the psychophysical data from Kröse (1986), and the results from Malik and Perona (1990). Here the data from Kröse (1986) was based on the converted data given in Malik and Perona (1990). Fig. 7 shows the data points linearly scaled so that the measures for the second pair (+ [ ]) match. Our measure predicts that that (+ O) is much easier to discriminate than all the other pairs, the pair (L_L, M_L) is barely discriminable with a score of 0.001, and the pair (R-mirror-R) is not discriminable with a score of −0.01.

It is clear from Table 1 that our model performance is entirely consistent with the other two. Note that the Malik and Perona model (Malik & Perona, 1990) is a representative of FRF models; thus our qualitative comparisons in Section 4.3 apply to their model. In addition, we employ only five commonly used filters instead of 96 pairs of filters in their model. Their model needs an elaborate form of nonlinearity that depends on

Table 1
Texture discrimination scores

Texture pair	Texture discriminability		
	Human data (Kröse, 1986)	Malik and Perona results (Malik & Perona, 1990)	Spectral histogram results
(+ O)	100	407	0.135
(+ [ ])	88.1	225	0.036
(L +)	68.6	203	0.027
(L M)	n.a.	165	0.023
(Δ →)	52.3	159	0.018
(+ T)	37.6	120	0.015
(+ X)	30.3	104	0.014
(T L)	30.6	90	0.004
(L_L, M_L)	n.a.	85	0.001
(R-mirror-R)	n.a.	50	−0.01

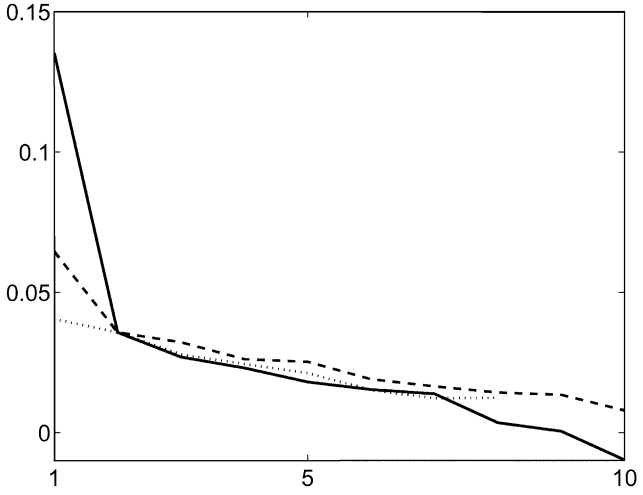

Fig. 7. Texture discrimination results. Here the horizontal axis corresponds to the order of the texture pairs in Table 1 and the vertical axis the texture discrimination scores. (...) Psychophysical data from Kröse (1986); (– – –) results from Malik and Perona's model (Malik & Perona, 1990); (——) results from the spectral histogram model.

inter-filter interactions specific to different filter types (they reported that simplified versions of this nonlinearity produce inferior performances).

As alluded to earlier, nonlinearity is an important property of human texture discrimination. Texture pairs (L, M) and (L_L, M_L) were constructed by Williams and Julesz (1992a) to argue against linear models. The (L, M) pair is among the ones that are easily discriminable. However, the (L_L, M_L) pair, which was constructed by simply adding a uniform texture of little L's at the endpoints of the L's and M's in the (L, M) pair, is not. This demonstrates that texture discrimination cannot be a simple linear operation; some form of nonlinearity must be included in order to account for this phenomenon. The reason is the following. The discriminability of the uniform texture of L's is zero. When this texture is added to the easily discriminable (L, M) texture, the discriminability of the resulting (L_L, M_L) texture should

equal that of the (L, M) texture if texture discrimination were a linear operation. But it is not: the discriminability of the (L_L, M_L) texture is in fact much lower (see Fig. 5). The Malik and Perona model (Malik & Perona, 1990) is able to reproduce this nonlinearity by incorporating two nonlinear stages. In contrast, our model reproduces the nonlinearity without any additional nonlinear operation.

According to Malik and Perona (1990), their model cannot account for asymmetry in texture discrimination, which refers to the phenomenon that one texture embedded in another one is more readily discriminated than when the latter is embedded in the former (Gurnsey & Browse, 1987; Williams & Julesz, 1992b). Our model may be able to account for the asymmetry phenomenon, and we illustrate this by a simulation involving the commonly used textures of +'s and L's. Fig. 8 shows test patterns used in our simulation. To be consistent with our previous evaluation methodology, we place one texture in the middle and the other one on the two sides. To reflect that the middle one forms the foreground, we compare the peak corresponding to a boundary separating two textures with the peak within the side (background) texture. Note that the layout in Fig. 8 yields two such scores, and the average is taken to indicate the discrimination strength.

For Fig. 8a, the discrimination score produced by our model is 0.005, and for Fig. 8b it is 0.018. In other words, our model predicts that the texture of +'s in the middle of the texture of L's is more difficult to discriminate than when the latter is in the middle of the former, hence discrimination asymmetry between the two textures. This prediction matches the psychological data (Gurnsey & Browse, 1987). The reason for our model to exhibit discrimination asymmetry is that the variability within the texture of L's is larger than that within the texture of +'s. As a result, the boundary separating the two textures can be relatively stronger or weaker compared to spurious boundaries generated within a background texture. This explanation is similar to that given by Rubenstein and Sagi (1990), who did a more systematic study on discrimination asymmetry based on an FRF model.

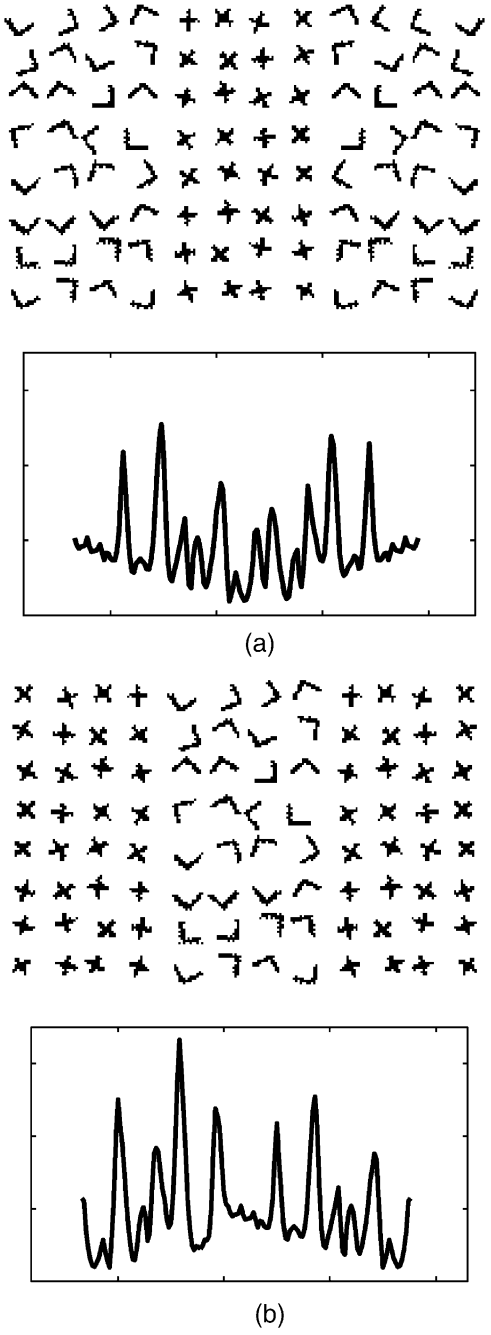

Fig. 8. Asymmetry in texture discrimination. (a) A texture region of +'s flanked by those of L's with the average texture gradient. The discrimination score produced by the spectral histogram model is 0.005 and the size of the image is 154×230. (b) A region of L's flanked by those of +'s. The discrimination score is 0.018 and the size of the image is 154×223.

4. Relation to other studies

This section clarifies the similarity and difference between the spectral histogram model and other related studies on texture modeling and texture discrimination. We first point out the relationship with the texton theory, and then relate to the texture processing literature that also employs some form of histogram analysis. Finally, we compare with FRF models for texture discrimination.

4.1. Relation to Julesz's texton theory

We are directly inspired by the texton theory for our employment and analysis of histograms. According to the texton theory (Julesz, 1981, 1995), preattentive texture segregation occurs between two regions only if they differ in texton density, irrespective of the spatial

relationships among textons. A histogram, or marginal distribution, provides a means to represent densities. However, our model builds on spatial/frequency filters and their responses are systematically represented in the spectral histogram. As a result, our model does not need to invoke specialized detectors, which would require a relatively late stage of perceptual computing as well as to limit the scope of the theory (see Section 2.2). Therefore, our model can be applied not only to synthetic textures commonly used in psychophysical experiments but also to natural textures, a point further discussed in Section 4.3.

4.2. Relation to other histogram-related studies

A histogram analysis is frequently used both for data analysis and image processing (Haralick & Shapiro, 1985). In the literature of texture processing, Unser (1986) used the sum and difference histograms of pixel pairs for texture classification. Voorhees and Poggio (1988) proposed to employ histograms to compare local distributions of blob detector responses in order to derive texture boundaries. To compare two histograms, they compute the maximum difference between corresponding histogram bins. We have compared their statistic and ours on a systematic database for texture classification, and found that the χ^2-statistic yields better performance (Liu & Wang, 2000). More comparisons with their model are given in Section 5.2. Based on responses from a nonlinear filter, Räth and Morfill (1997) used a histogram comparison to derive texture boundaries. Ojala, Pietikainen, and Harwood (1996) and Hofmann, Puzicha, and Buhmann (1998) also employed some response histograms for texture processing. In contrast to these studies, which treat a histogram as a texture feature for producing good empirical results, we treat the spectral histogram as an explicit model of texture and verify its validity using texture synthesis.

Our model can be treated as an extension to the Heeger and Bergen model proposed for texture synthesis (Heeger & Bergen, 1995). The main difference lies in the synthesis procedure; ours uses statistical sampling that is guided directly by the histogram difference between an original image and a synthesized one, whereas theirs matches histograms independently in an image pyramid. Their iterative algorithm is computationally efficient, but does not guarantee convergence; they reported that it generally converges in a few iterations (see also Portilla & Simoncelli, 2000). More problematic is that after convergence there is no assurance that the histogram of the reconstructed image is close to that of the original one. Our simulations with their algorithm indicate that the algorithm is particularly prone to the local minimum problem for synthetic textures studied in this paper; that is, the algorithm converges to an image whose histogram is quite different from that of the original. This is illustrated in Fig. 9 with the textures of circles and pluses used in Fig. 3. The middle image in Fig. 9a shows the synthesized texture by the Heeger and Bergen algorithm when the original image is the left one in Fig. 9a. The quality of synthesis is reasonable but not as good as ours in Fig. 4c. When the original texture is the left one in Fig. 9b, the quality of their synthesized texture given in the middle of Fig. 9b becomes worse. To investigate whether the lower quality is caused by the filters used or the synthesis procedure, the right images in Fig. 9a and b show the corresponding results from our algorithm using the same steerable filters employed in their model. The quality of synthesis is significantly improved. Quantitatively, the spectral histogram difference between the synthesized texture and original one in Fig. 9a is 0.111 per filter for their algorithm and it is reduced to 0.016 for our algorithm. For Fig. 9b, the difference for their algorithm is 0.326 per filter and it is reduced to 0.013 for our algorithm. This clearly indicates that their synthesis procedure is a main cause for the relatively poor performance.

Zhu et al. (1997) studied texture synthesis by learning a probability model using histograms of filter responses. First the probability model is learned based on the observed image(s). Then the texture synthesis is achieved by sampling the learned probability model. While the system is successful in synthesizing natural textures, the learned probability model seems ineffective for synthetic textures; even with texture elements used as filters directly and a specially designed sampling procedure, the synthesized textures are perceptually different from the observed ones (Zhu et al., 1997). In contrast, our texture model is defined by spectral histograms, and it is conceptually consistent with the texton theory. Our improved sampling procedure makes it possible to synthesize challenging textures used in psychophysical experiments.

4.3. Comparison with FRF models

Essentially our model consists of three stages: a filtering stage, a histogram gathering stage, and a histogram comparison stage. In comparison with FRF models, ours is a filter-histogram-contrast model. Our first stage is the same as in FRF models. We do not need a rectifying nonlinearity because a spectral histogram reflects statistics higher than the first-order moment (mean). To explain this point, we show the histogram responses to two images of identical mean but different variances. Such examples are commonly used to justify the rectifying nonlinearity. Fig. 10 shows an image with the spectral histograms of the left and right half. The image was generated by adding Gaussian noise with different variances to a uniform image and thus the left and right regions have identical mean but different variances. However, their spectral histograms are very

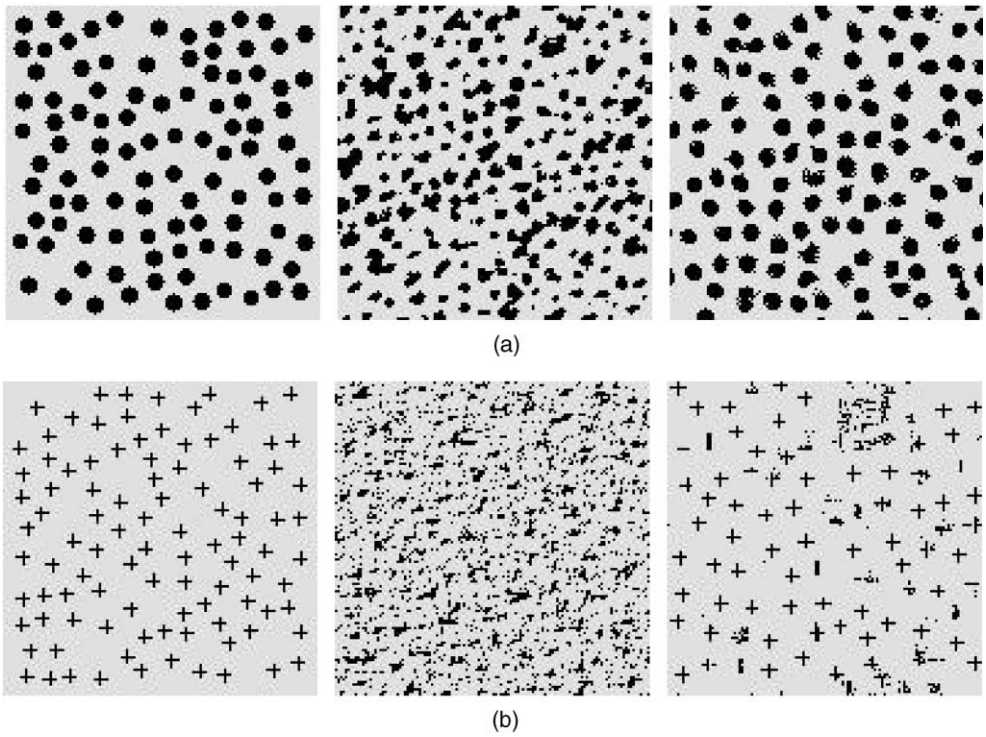

Fig. 9. Comparison with the Heeger and Bergen algorithm (Heeger & Bergen, 1995) for synthetic texture synthesis. In each row, the left column shows the observed image, the middle a synthesized texture using their algorithm, and the right a synthesized texture using the sampling algorithm given in Appendix A. Here the same steerable filters are used in both synthesis algorithms. (a) A texture consisting of circles. The difference between the observed histogram and the synthesized one is 0.111 per filter for their algorithm and 0.016 for our algorithm. (b) A texture consisting of pluses. The difference between the observed histogram and the synthesized one is 0.326 per filter for their algorithm and 0.013 for our algorithm.

different in the relative heights of peaks, resulting in a large χ^2-distance between them. After rectification, a pooling operation is needed in FRF models for reducing inhomogeneity in filter responses. Because a histogram is gathered from a window of a particular integration scale, it implicitly performs a kind of spatial smoothing. The third stage for computing a texture boundary is similar between our model and FRF models, which involves a difference (or similarity) measure. In our case, because histograms are marginal distributions, we use χ^2-statistic to measure the difference between histograms (other statistics can also be used, see Liu, 1999). For FRF models, the comparison is between two spectra of various filter responses.

The above discussion makes it clear that our model is simpler at the conceptual level. It does not need a rectifying nonlinearity and subsequent pooling; the latter has been argued to require another nonlinear operation (Malik & Perona, 1990). The computational functions of such operations are intrinsically incorporated in histogram gathering. Thus, our model is more parsimonious.

Because the visual system normally deals with natural images, a good texture model should, in addition to discriminating synthetic textures, perform well on classifying real textures. Rarely are standard models evaluated on real textures. A main reason for this is that psychophysical experiments almost always use binary, synthetic textures. Such impoverished stimuli are often necessary in controlled human experiments. However, one undesirable consequence is that resulting theories and computational models are often limited to just such stimuli, not applicable to natural images. We think that popular notions such as line, corner, and terminator detectors, in the texture perception literature, even the texton theory itself, are colored by the use of laboratory stimuli.

Randen and Husoy (1999) recently performed an extensive evaluation of various texture classification methods that are based on filter responses directly. Their system setup for comparing purposes includes filtering, nonlinearity, smoothing, and then classification. Thus, the setup can be viewed as an FRF model. Their comparisons conclude that no method performs consistently well on natural images, and this comprehensive study suggests that FRF models are inadequate to classify natural textures (see also Chubb & Landy, 1991). On the other hand, the spectral histogram has been successfully used to classify a large number of real texture images (Liu, 1999; Liu & Wang, 2000). Our systematic comparison shows that the spectral histogram model substantially outperforms FRF models.

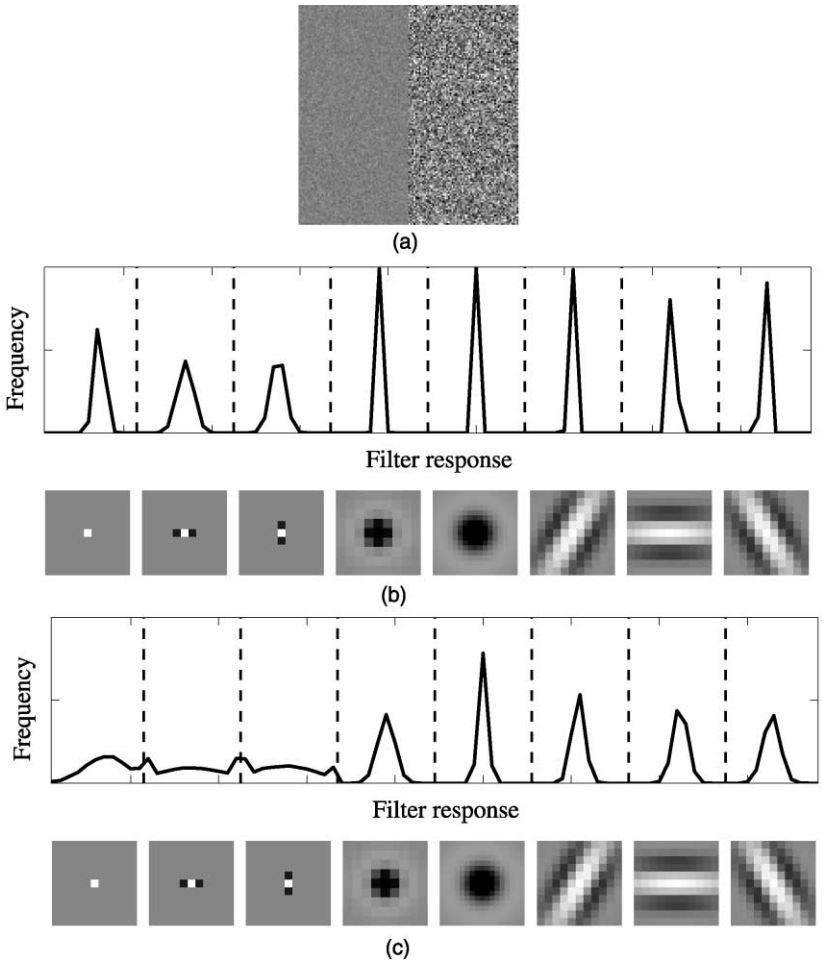

Fig. 10. The spectral histograms of two regions with an identical mean but different variances. Here the same eight filters as in Fig. 1 are used for illustration. (a) An image consisting of such two regions, which is generated by adding Gaussian noise with different variances to a uniform image. The left region has a variance of 10 and the right 50. The size of the image is 128×128. (b) The spectral histogram of the left region. (c) The spectral histogram of the right region.

Fig. 10 illustrates that a histogram can discriminate the second-order moment (variance), thus no need for rectification. Indeed, a spectral histogram encodes all the higher order moments. In contrast, FRF models can only discriminate differences in the second-order moment, because discrimination is essentially based on filter-response energy, which corresponds to a second-order statistic (for more discussions see Kingdom et al., 2001). Recently, Kingdom et al. (2001) tested the sensitivity of human observers to differences in histograms of synthetic wavelet-textures. Their experiments show that human subjects are sensitive to not only the second-order moment between two wavelet histograms but also the third-order (skew) and the fourth-order (kurtosis) moments. This result clearly implies that FRF models are inadequate for human texture discrimination. On the other hand, the spectral histogram model measures directly histogram differences, and it is sensitive to higher order moments such as skew and kurtosis. This analysis shows that the spectral histogram model is more general than an FRF model, and it is reduced to the latter when higher than second-order statistics are ignored from the histograms.

5. Discussion

5.1. Filter selection

The performance of all filter-based models, including ours, inevitably depends on the choice of filters. For example, if no color filter is used a model cannot characterize or discriminate color-defined textures. In the other extreme, as discussed in Section 2.1, given sufficient filters the spectral histogram model can uniquely represent an arbitrary image up to a translation. The model described here uses three common types of filter: LoG, Gabor, and difference. Our extensive synthesis results on both synthetic and natural textures show that a fixed set of filters is often sufficient to capture texture characteristics. The choice of specific parameter values for each filter is obviously motivated by performance

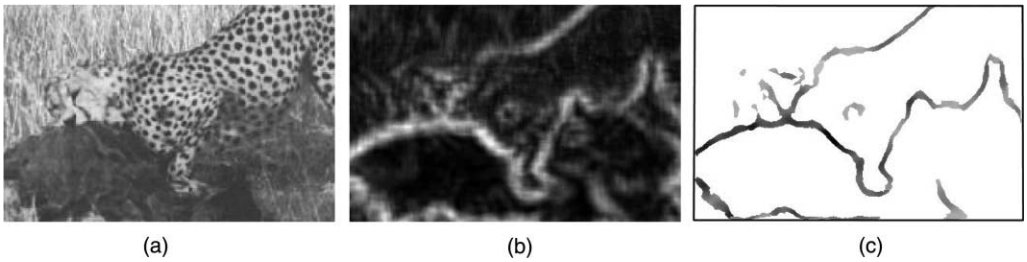

Fig. 11. Boundary detection for a natural texture image: (a) input image, whose size is 277×422; (b) texture gradient produced by the spectral histogram model; (c) detected texture boundaries.

considerations. However, it is not a difficult task given some general understanding about the filter and the texture.

Two topics related to filter selection have been studied before. The first is known as *filter design*: given a filter family, filter design aims to choose the best parameters for a given task. Filter design has been studied extensively for Gabor filters, and the most common method is to identify local peaks in the frequency domain and then align the filters around the identified peaks (Bovik, Clark, & Geisler, 1990; Geisler & Hamilton, 1986; Heeger, 1987). As shown in a comparative study on different kinds of filters (Randen & Husoy, 1999), no single filter type can give consistent results on several texture sets. The second topic addresses the selection of a subset from a large filter bank for a given task. Because it is computationally infeasible to exhaustively search for the best subset when the number of filters is large, greedy (locally optimal) algorithms are often adopted in practice, where the best filter together with the ones already chosen forms the next choice and the procedure is repeated until reaching some performance requirement (Campbell & Thomas, 1997; Jain & Farrokhnia, 1991; Liu & Wang, 2001; Zhu et al., 1997).

5.2. Texture segregation

The goal of texture segregation is to produce boundaries separating different texture regions, or, conversely, to segregate an image into regions of homogeneous texture. A potential issue with the spectral histogram model is that, because histogram gathering requires a sizable window, it may lead to grossly inaccurate boundaries. Note that this issue is not unique to our model, and standard models for texture segregation all include a stage of spatial pooling, which has an effect of blurring boundaries. In essence, texture features require a larger spatial scale to manifest than, say, luminance, color, or motion features.

To illustrate how our model can apply to segmenting natural textures, we process an image that was first used in Voorhees and Poggio (1988). The image, shown in Fig. 11a, contains a cheetah biting a buffalo. Fig. 11b shows the texture gradient of the image produced by our model. The gradient at a pixel location is the sum of the χ^2-distances between the spectral histograms of adjacent windows along a row and a column. In a row or column, the gradient is calculated in the same way as in our texture discrimination experiments. Given the gradient, we detect the texture boundaries by finding local maxima in the gradient image and the resulting texture boundaries are given in Fig. 11c. The cheetah boundaries in Fig. 11c are more extensive and accurate than that generated by Voorhees and Poggio (1988). Ours also yields the boundaries of the buffalo, while theirs does not because their system is specifically designed for blob-like textures.

Fig. 11 is meant to be an illustration. Segmentation of natural textures in the context of spectral histograms is a topic to be dealt with in a separate study (see also Liu, 1999), where we suggest a subsequent stage for accurate boundary localization.

5.3. Biological plausibility

In Section 4.3, we have discussed that our model, being sensitive to higher than second-order statistics, is consistent with human texture perception while FRF models are not adequate. Using a class of independent, identically distributed textures, Chubb et al. (1994) illustrated that histogram contrast might be used by the visual system to draw distinctions between different image regions. An analysis performed by Kingdom et al. (2001) suggests that a model based on response histograms from Gabor filters is superior to a model based on pixel histograms. These results are also consistent with our model and lend direct support to our use of filters and their response histograms. On the other hand, much work—both empirical and theoretical—remains to be done to characterize human sensitivity to histogram contrast.

One advantage of FRF models, e.g. the Malik and Perona model, is that its components are biologically plausible. How plausible is our filter-histogram-contrast model biologically? The filtering stage in our model is commonly used in other models of texture discrimination, and as previously mentioned early processing by spatial/frequency filters in the visual system is widely

accepted (Campbell & Robson, 1968; De Valois & De Valois, 1988). More specifically, our study has employed three types of filters: LoG filters, Gabor filters, and difference filters. Physiological evidence supports the existence of neurons whose response properties resemble LoG and Gabor filters. Difference filters correspond to simple edge detectors along the horizontal and the vertical direction (see Fig. 1), and this processing can be carried out by simple cells in the visual cortex (Hubel, 1988).

To compute a statistical quantity, such as a histogram bin, requires neurons with sizable receptive fields, which would presumably occur in the extrastriate cortex (Kandel, Schwartz, & Jessell, 1991; Schiller, 1995). Neurophysiological experiments found that many neurons in cortical area V2, but not in V1, respond to illusory contours (von der Heydt & Peterhans, 1989; von der Heydt, Peterhans, & Baumgartner, 1984). Furthermore, Merigan, Nealey, and Maunsell (1993) found that V2 lesions damage the ability of monkeys to discriminate the orientations of texture boundaries. These results point to a crucial role by V2 in texture discrimination (see also Wilson, 1999).

Little is known about how a visual neuron might calculate a statistic by pooling the responses from a population of other cells. Hence, our following discussion is inevitably speculative. A unique aspect of histogram gathering is that the entire response spectrum needs to be considered. Our simulations demonstrate, however, that a few histogram bins are often sufficient for the purpose of texture synthesis and discrimination. Let us consider how a single bin could be calculated. To detect whether the response of a linear filter (presumably in V1) falls into the bin needs two conditions to be simultaneously satisfied: the response needs to be greater than one threshold and less than another. So, this comes down to a conjunction of two thresholding operations, which is easy to implement even with the simplest kind of a neuron model: the McCulloch–Pitts model (McCulloch & Pitts, 1943). Once this is computed, the histogram bin would simply be a summation from a group of such conjunction neurons. Multiple bins of the same filter as well as multiple filters would be calculated in the similar fashion.

Given coded histogram bins, a contrast measure between them, such as the χ^2-statistic in (2), can be first decomposed into local comparisons within individual bins. A subsequent operation involving spatial summation and shunting inhibition would yield the outcome.

6. Conclusion

We have suggested the spectral histogram model as a quantitative definition for texton patterns, and demonstrated that model predictions well match a systematic set of texture discrimination data. In addition, the model exhibits both nonlinearity and asymmetry in texture perception. While the present study focuses on textures commonly used in psychophysical experiments, the spectral histogram model performs equally well on natural textures.

In our view, the lack of a quantitative formulation for the texton theory becomes a major obstacle to progress, and our model represents a first attempt towards this end. As such, our model is almost certainly wrong in many aspects. On the other hand, our model provides a solution to several key issues in characterizing elusive textons and in modeling texture discrimination.

Acknowledgements

We thank two anonymous reviewers whose insightful comments have significantly improved the presentation of this paper and T.F. El-Maraghi for making his implementation of the Heeger and Bergen synthesis algorithm available online. The research described here was supported in part by a NIMA grant (NMA201-01-1-2010) and an FYAP grant from the Florida State University to XL, and an ONR Young Investigator Award (N00014-96-1-0676), an NSF grant (IIS-0081058), and an AFOSR grant (F49620-01-1-0027) to DLW.

Appendix A. Sampling algorithm for binary textures

Here we give the sampling algorithm used in this paper for completeness. The algorithm was proposed by Zhu et al. (1997) for learning probability models of images. Before using this algorithm, one needs to specify the filters ($F^{(\alpha)}$) to be used, the number of histogram bins ($L^{(\alpha)}$) and the bin locations for each filter, τ, which controls the learning speed, and ϵ, which specifies the stopping criterion. The histogram ($H^{(\alpha)}_{\mathbf{I}}$) is obtained by first convolving $\mathbf{I}$ with $F^{(\alpha)}$ and then computing the histogram of the filtered image. For all the synthesis results in this paper, we have used 47 filters (see Section 2.2) including local difference filters, LoG filters, and Gabor filters. We have used three bins for the local difference filters and 11 bins for all other filters; other values can also be used. The bin locations are determined by dividing the filter-response range into the specified number of bins of equal size. In our experiments, τ is fixed to 0.035 and ϵ is 0.025. $\lambda^{(\alpha)}_i$, the parameter associated with the corresponding histogram bin, determines its influence on P_{black} and is learned along with the sampling process. While the algorithm can be implemented straightforwardly, an efficient implementation requires the effective computation of $H^{(\alpha)}_{\mathbf{I}_{\text{black}}}$ and $H^{(\alpha)}_{\mathbf{I}_{\text{white}}}$ from $H^{(\alpha)}_{\mathbf{I}_{\text{syn}}}$ based on the following fact: one of

$\mathbf{I}_{\text{black}}$ and $\mathbf{I}_{\text{white}}$ is the same as $\mathbf{I}_{\text{syn}}$ and the other differs from $\mathbf{I}_{\text{syn}}$ by only one pixel.

The synthesis algorithm is as follows:

For a binary input texture, compute $H^{(\alpha)}_{\text{obs}}$, $\alpha = 1, \ldots, K$.
Initialize $\mathbf{I}_{\text{syn}}$ as a binary white noise image and $\lambda^{(\alpha)}_i \leftarrow 0$.
  Repeat
    For each pixel location $\vec{v}$ in $\mathbf{I}_{\text{syn}}$, do
      $\mathbf{I}_{\text{black}} \leftarrow \mathbf{I}_{\text{syn}}$, $\mathbf{I}_{\text{black}}(\vec{v}) \leftarrow 0$, $\mathbf{I}_{\text{white}} \leftarrow \mathbf{I}_{\text{syn}}$, $\mathbf{I}_{\text{white}}(\vec{v}) \leftarrow 1$.
      Compute $H^{(\alpha)}_{\mathbf{I}_{\text{black}}}$ and $H^{(\alpha)}_{\mathbf{I}_{\text{white}}}$, $\alpha = 1, \ldots, K$.
      $E_{\text{black}} \leftarrow \sum_{\alpha=1}^{K} \sum_{i=1}^{L^{(\alpha)}} \lambda^{(\alpha)}_i \times H^{(\alpha)}_{\mathbf{I}_{\text{black}}}(i)$,
      $E_{\text{white}} \leftarrow \sum_{\alpha=1}^{K} \sum_{i=1}^{L^{(\alpha)}} \lambda^{(\alpha)}_i \times H^{(\alpha)}_{\mathbf{I}_{\text{white}}}(i)$
      $P_{\text{black}} \leftarrow \exp(-E_{\text{black}}) / (\exp(-E_{\text{black}}) + \exp(-E_{\text{white}}))$.
      $\mathbf{I}_{\text{syn}}(\vec{v}) \leftarrow 0$ with probability P_{black} and $\mathbf{I}_{\text{syn}}(\vec{v}) \leftarrow 1$ with $1 - P_{\text{black}}$.
    $\lambda^{(\alpha)}_i \leftarrow \lambda^{(\alpha)}_i + \tau(H^{(\alpha)}_{\mathbf{I}_{\text{syn}}}(i) - H^{(\alpha)}_{\text{obs}}(i))$.
  Until $\sum_{i=1}^{L^{(\alpha)}} |H^{(\alpha)}_{\mathbf{I}_{\text{syn}}}(i) - H^{(\alpha)}_{\text{obs}}(i)| \leqslant \epsilon$ for $\alpha = 1, 2, \ldots, K$.

References

Barth, E., Zetzsche, C., & Rentschler, I. (1998). Intrinsic two-dimensional features as textons. *Journal of the Optical Society of America A, 15*, 1723–1732.

Beck, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science, 154*, 538–540.

Beck, J. (1982). Textural segmentation. In J. Beck (Ed.), *Organization and representation in perception* (pp. 285–318). Hillsdale, NJ: Erlbaum.

Bergen, J. R. (1991). Theories of visual texture perception. In D. Regan (Ed.), *Spatial vision* (pp. 114–134). Boca Raton: CRC Press.

Bergen, J. R., & Adelson, E. H. (1988). Early vision and texture perception. *Nature, 333*, 363–367.

Bovik, A. C., Clark, M., & Geisler, W. S. (1990). Multichannel texture analysis using localized spatial filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 12*, 55–73.

Caelli, T. M. (1985). Three processing characteristics of visual texture segmentation. *Spatial Vision, 1*, 19–30.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology (London), 197*, 551–566.

Campbell, N. W., & Thomas, B. T. (1997). Automatic selection of Gabor filters for pixel classification. In *Proceedings of the sixth international conference image processing and its applications* (pp. 761–765).

Chubb, C., Econopouly, J., & Landy, M. S. (1994). Histogram contrast analysis and the visual segregation of IID textures. *Journal of the Optical Society of America A, 11*, 2350–2374.

Chubb, C., & Landy, M. S. (1991). Orthogonal distribution analysis: A new approach to the study of texture perception. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 291–301). Cambridge, MA: MIT Press.

Cross, G. R., & Jain, A. K. (1983). Markov random field texture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 5*, 25–39.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.

Dubes, R. C., & Jain, A. K. (1993). Random field models in image analysis. *Journal of Applied Statistics, 16*, 131–164.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification*. New York: Wiley.

Fogel, I., & Sagi, D. (1989). Gabor filters as texture discriminators. *Biological Cybernetics, 61*, 103–113.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 721–741.

Geisler, W. S., & Hamilton, D. B. (1986). Sampling theory analysis of spatial vision. *Journal of the Optical Society of America A, 3*, 62–70.

Graham, N., Beck, J., & Sutter, A. (1992). Nonlinear processes in spatial-frequency channel models of perceived texture segregation: Effects of sign and amount of contrast. *Vision Research, 32*, 719–743.

Gurnsey, R., & Browse, R. A. (1987). Micropattern properties and presentation conditions influencing visual texture discrimination. *Perception and Psychophysics, 41*, 235–252.

Haralick, R. M., & Shapiro, L. G. (1985). Image segmentation techniques. *Computer Graphics and Image Processing, 29*, 100–132.

Heeger, D. J. (1987). Optical flow from spatiotemporal filters. In *Proceedings of the first international conference on computer vision* (pp. 181–190).

Heeger, D. J., & Bergen, J. R. (1995). Pyramid-based texture analysis/synthesis. In *Proceedings of SIGGRAPHS* (pp. 229–238).

Hofmann, T., Puzicha, J., & Buhmann, J. M. (1998). Unsupervised texture segmentation in a deterministic annealing framework. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 803–818.

Hubel, D. H. (1988). *Eye, brain, and vision*. New York: Freeman & Co.

Jain, A. K., & Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern Recognition, 23*, 1167–1186.

Julesz, B. (1962). Visual pattern discrimination. *IRE Transactions on Information Theory, 8*, 84–92.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290*, 91–97.

Julesz, B. (1986). Texton gradients: The texton theory revisited. *Biological Cybernetics, 54*, 245–251.

Julesz, B. (1995). *Dialogues on perception*. Cambridge, MA: MIT Press.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (1991). *Principles of neural science* (3rd ed.). New York: Elsevier.

Kingdom, F. A. A., Hayes, A., & Field, D. J. (2001). Sensitivity to contrast histogram differences in synthetic wavelet-textures. *Vision Research, 41*, 585–598.

Kröse, B. J. (1986). A description of visual structure. Ph.D. Dissertation, Delft University of Technology, The Netherlands.

Liu, X. (1999). Computational investigation of feature extraction and image organization. Ph.D. Dissertation, The Ohio State University, Columbus, OH. Available: http://fsvision.fsu.edu/publications/papers/dissertation/index.html.

Liu, X., & Wang, D. L. (2000). Texture classification using local spectral histograms. Technical Report OSU-CISRC-7/00-TR17, Department of Computer and Information Science, The Ohio State University, Columbus, OH. Available: ftp://ftp.cis.ohio-state.edu/pub/tech-report/2000/TR17.ps.gz.

Liu, X., & Wang, D. L. (2001). Appearance-based recognition using perceptual components. *Proceedings of the international joint conference on neural networks* (vol. 3, pp. 1943–1948).

Malik, J., & Perona, P. (1990). Preattentive texture discrimination with early vision mechanisms. *Journal of the Optical Society of America A, 7*, 923–932.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115–133.

Merigan, W. H., Nealey, T. A., & Maunsell, J. H. R. (1993). Visual effects of lesions of cortical area V2 in macaques. *Journal of Neuroscience, 13*, 3180–3191.

Ojala, T., Pietikainen, M., & Harwood, D. (1996). A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition, 29*, 51–59.

Papathomas, T. V., Chubb, C., Gorea, A., & Kowle, E. (Eds.). (1995). *Early vision and beyond*. Cambridge, MA: MIT Press.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40*, 49–71.

Randen, T., & Husoy, J. H. (1999). Filtering for texture classification: A comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*, 291–310.

Räth, C., & Morfill, G. (1997). Texture detection and texture discrimination with anisotropic scaling indices. *Journal of the Optical Society of America A, 14*, 3208–3215.

Rubenstein, B. S., & Sagi, D. (1990). Spatial variability as a limiting factor in texture-discrimination tasks: Implications for performance asymmetries. *Journal of the Optical Society of America A, 7*, 1632–1643.

Schiller, P. H. (1995). Visual processing in the primate extrastriate cortex. In T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 167–176). Cambridge, MA: MIT Press.

Simoncelli, E. P., Freeman, W. T., Adelson, E. H., & Heeger, D. J. (1992). Shiftable multi-scale transforms. *Wavelets [special issue]. IEEE Transactions on Information Theory, 38*, 587–607.

Turner, M. R. (1986). Texture discrimination by Gabor functions. *Biological Cybernetics, 55*, 71–82.

Unser, M. (1986). Sum and difference histograms for texture classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 8*, 118–125.

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex: I. Lines of pattern discontinuity. *Journal of Neuroscience, 9*, 1731–1748.

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*, 1260–1262.

Voorhees, H., & Poggio, T. (1988). Computing texture boundaries from images. *Nature, 333*, 364–367.

Watt, R. J. (1995). Some speculations on the role of texture processing in visual perception. In T. V. Papathomas, C. Chubb, A. Gorea, & E. Kowler (Eds.), *Early vision and beyond* (pp. 59–67). Cambridge, MA: MIT Press.

Williams, D., & Julesz, B. (1992a). Filters versus textons in human and machine texture discrimination. In H. Wechsler (Ed.), *Human and machine perception* (pp. 145–175). San Diego, CA: Academic Press.

Williams, D., & Julesz, B. (1992b). Perceptual asymmetry in texture perception. *Proceedings of the National Academy of Science USA, 89*, 6531–6534.

Wilson, H. R. (1999). Non-Fourier cortical processes in texture, form, and motion perception. *Cerebral Cortex, 13*, 445–477.

Zhu, S. C., Liu, X., & Wu, Y. N. (2000). Exploring texture ensembles by efficient Markov chain Monte Carlo. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22*, 554–569.

Zhu, S. C., Wu, Y. N., & Mumford, D. (1997). Minimax entropy principle and its application to texture modeling. *Neural Computation, 9*, 1627–1660.